Least-Squares

Luke Vilaseca

December 2018

1 Introduction

Regression is the processes for estimating the relationship between variables. The main purpose of regression is to examine if a set of predictor variables do a well enough job at predicting an outcome variable, known as a dependent variable. It also serves the purpose as to figure out which variables in particular are significant predictors of the outcome variable. This relationship between two independent variables and the dependent outcome are explained by a simple regression formula equal to

$$y = \alpha(x) + \beta + \epsilon,$$

The $\alpha(x)$ serves the purpose of giving the linear regression line its slope. The slope is a best fit line of a given data set. More often than not, the slope given to a regression line does not go through every point. Therefore, there is always an error that must be accounted for. This type of error or average distance from a point to the line is given by the ϵ. In addition, β is a point of reference for the data set to begin. This β is the y-intercept and is the jumping off point for a regression line.

2 Regression Analysis

Regression analysis is a mathematical way of determining the impact of some selected variables. These variables follow the trend of one or more being independent and the other dependent. The difference between the two is that the independent variables are suspected of having an impact on the dependent variable. Therefore, you are trying to predict or understand what the dependent variable actually is.

To conduct a regression analysis, one must first gather data on the variables you wish to find a correlation between and plot the information on a graph. For instance, consider the data displayed in Figure 1. Here, the correlation between independent variable, height, and dependent, weight.

The regression line is using the height of an individual to predict their weight. This linear formula for determining your dependent variable is displayed as

$$y = mx + b.$$

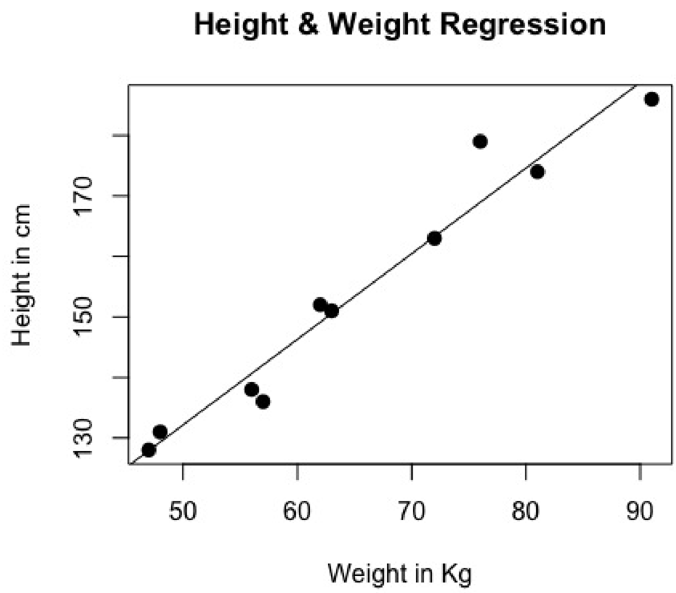

Figure 1: Scatter plot with a linear regression line showing the correlation.

As seen here, regression can be used to determine a probable mean for a given x. Regression analysis is an exploration to understand the behavior of a given independent variable and a predicted one as well as gives information on the significance of the relationship.

3 Linear Algebra

From a linear algebra point of view, regression cannot simply be found by using a $A\vec{x} = \vec{b}$ equation. This is due to the fact that the $A\vec{x} = \vec{b}$ will come out to be inconsistent, meaning it has no solution. Take this following example of the $A\vec{x} = \vec{b}$ equation failing to produce a best fit line.

Example 1. Find a line that matches the given data set.

X	Y
1	2
2	6
4	1

Set up a linear system using the $y = mx + b$ equation.

$$\begin{aligned} 2 &= m1 + b \\ 6 &= m2 + b \\ 1 &= m4 + b \end{aligned}$$

Now put in the form of $A\vec{x} = \vec{b}$, where

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 1 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} m \\ b \end{pmatrix}, \quad \text{and} \quad \vec{b} = \begin{pmatrix} 2 \\ 6 \\ 1 \end{pmatrix}.$$

Set up the augmented matrix and row-reduce,

$$\left(\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & 1 & 6 \\ 4 & 1 & 1 \end{array}\right) \xrightarrow{RR} \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & -5 \end{array}\right).$$

After row reduction, we see that $A\vec{x} = \vec{b}$ is inconsistent. A matrix is inconsistent when there is no solution $\vec{x}$ to the equation $A\vec{x} = \vec{b}$, which is evident in this row reduction by the third row. The third row of the matrix in Example 1 is saying that $0 = -5$ which is not possible. Therefore, there is line that can describe all the points because $A\vec{x} = \vec{b}$ is inconsistent.

However, we can use $A\vec{x}$ as an approximation of $\vec{b}$ minimizing the smaller the distance between the $A\vec{x}$ and $\vec{b}$, yields a better approximation. Finding the vector $\vec{x}$ that yields the smallest value of $\|\vec{b} - A\vec{x}\|$ is know as the *general least-squares problem.*

Definition 1. If A is $m \times n$ and $\vec{b}$ is in $\mathbb{R}^m$, a *least-squares solution* of $A\vec{x} = \vec{b}$ is an $\hat{x}$ in $\mathbb{R}^n$ such that

$$\|\vec{b} - A\hat{x}\| \leq \|\vec{b} - A\vec{x}\|,$$

for all $\vec{x}$ in $\mathbb{R}^n$.

Definition 2. The *column space* of an $m \times n$ matrix A, written as Col A, is the set of all linear combinations of the columns of A. If $A = [\vec{a_1}, \ldots, \vec{a_n}]$, then

$$\text{Col}\, A = \text{Span}(\vec{a_1}, \ldots, \vec{a_n}).$$

In essence, we are looking for an $\vec{x}$ that makes $A\vec{x}$ closest to the vector$\vec{b}$. In, particular, we seek $\vec{x}$ such that $A\vec{x}$ is the orthogonal projection of $\vec{b}$ onto the column space of A. This is due to the fact that the Col A is the range of linear transformation from $\vec{x}$ to $A\vec{x}$. To picture this, imagine a right triangle perpendicular to column space A where the estimated $\vec{b}$, $\hat{b}$, is the hypotenuse and the $\vec{b}$ is the side adjacent to the hypotenuse which lies in the column space of A as shown in Figure 2. This approximation can be described by the normal equation

$$A^T A\vec{x} = A^T\vec{b}$$

for $A\vec{x} = b$ for $A\vec{x} = \vec{b}$.

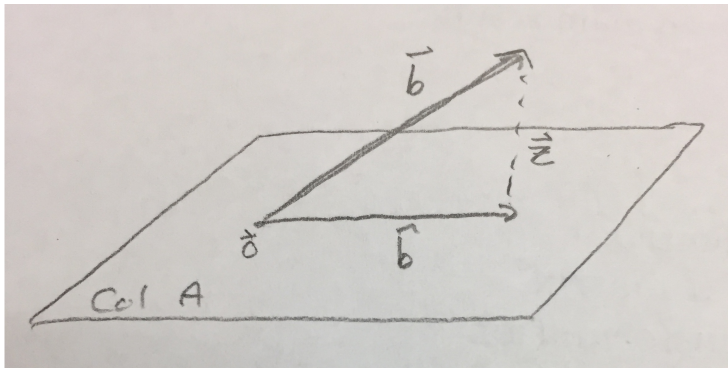

Figure 2: The Orthogonal Project onto $ColA$.

Theorem 3. *The set of least-squares solutions of the equation, $A\vec{x} = \vec{b}$, coincides with the nonempty set of solutions of the normal equations $A^T A\vec{x} = A^T\vec{b}$.*

To fully understand the normal equation, it is prudent to state the *Orthogonal Decomposition Theorem.*

Theorem 4. *Let W be a subspace of $\mathbb{R}^n$. Then each $\vec{b}$ in $\mathbb{R}^n$ can be written uniquely in the form*

$$\vec{b} = \hat{b} + \vec{z},$$

where $\hat{y}$ is in W and $\vec{z}$ is in the subspace perpendicular to W. The vectors can be rewritten to be

$$\vec{z} = \vec{b} - \hat{b}.$$

Proof. Recall: the least square solution to $A\vec{x} = \vec{b}$ can be computed by the normal equation, $A^T A\vec{x} = A^T\vec{b}$.

Recall: By the *Orthogonal Decomposition Theorem*, we have $\vec{b} = \hat{b} + \vec{z}$. Hence, we have $\vec{z} = \vec{b} - \hat{b}$.

We want to show that every least-squares solution, $\hat{x}$, to the equation, $A\vec{x} = \vec{b}$, is a solution to the normal equation $A^T A\vec{x} = A^T\vec{b}$.

In order to show this, we most show that every least squares solution, $\hat{x}$, satisfies the normal equation, $A^T A\vec{x} = A^T\vec{b}$. In addition, we must show that every solution to the normal equation is a least squares solution.

We start with showing that every least squares solution, $\hat{x}$, satisfies the normal equation, $A^T A\vec{x} = A^T\vec{b}$.

Decompose $\vec{b}$ as $\vec{b} = \hat{b}+\vec{z}$ as given in the *Orthogonal Decomposition Theorem.* Then $\vec{z}$ is perpendicular to the column space of A (2. Since $\hat{b}$ is in $ColA$, the equation $A\vec{x} = \hat{b}$ has at least one solution, $\hat{x}$. Therefore, $\hat{b} = A\hat{x}$. With this bit of information, we substitute $\hat{b} = A\hat{x}$ into the *Orthogonal Projection Theorem* equation. So you get $\vec{z} = \vec{b} - A\hat{x}$.

Let

$$A^T = \begin{pmatrix} \vec{a_1} \\ \vdots \\ \vec{a_n} \end{pmatrix}.$$

Since A^T is in the $ColA$, we take $A^T\vec{z}$ and get

$$A^T\vec{z} = \begin{pmatrix} \vec{a_1} \\ \vdots \\ \vec{a_n} \end{pmatrix} \vec{z} = \begin{pmatrix} \vec{a_1} \cdot \vec{z} \\ \vdots \\ \vec{a_n} \cdot \vec{z} \end{pmatrix} = \vec{0}.$$

We assumed earlier that $\vec{z}$ and $\operatorname{Col} A$ are perpendicular. Therefore, the dot product of the rows of A^T and $\vec{z}$ is $\vec{0}$. Since $A^T\vec{z} = \vec{0}$, we get

$$A^T\vec{z} = \vec{0}$$

$$A^T(\vec{b} - A\hat{x}) = \vec{0}$$

due to the fact that $\vec{z} = \vec{b} - A\hat{x}$. Since matrices are distributable by the matrix distribution property, we get

$$A^T\vec{b} - A^TA\hat{x} = \vec{0}$$

$$A^TA\hat{x} = A^T\vec{b}.$$

Therefore, every least squares solution, $\hat{x}$, satisfies the normal equation. Next, we want to show that every solution to the normal equation is a least square solution.

Suppose that we have a solution, $\hat{x}$, to the normal equation. Then

$$A^TA\hat{x} = A^T\vec{b}.$$

If we use the *Orthogonal Projection Theorem* equation and substitute $\vec{z} = \vec{b} - A\hat{x}$ and $\hat{b}$ for $A\hat{b}$

$$\vec{b} = A\hat{x} + (\vec{b} - A\hat{x}),$$

then we need to show that $\vec{b} - A\hat{x}$ is perpendicular to $ColA$.

$$A^T(\vec{b} - A\hat{x}) = A^T\vec{b} - A^TA\hat{x} = \vec{0}.$$

We know this because earlier we asserted that $A^TA\vec{x} = A^T\vec{b}$. Therefore, the difference of the two vectors gives you the $\vec{0}$. Therefore, by the *Orthogonal Decomposition Theorem*, every solution to the normal equation is a least square solution.

Therefore, there exists a solution, $\hat{x}$, to the least squares equation, $A\vec{x} = \vec{b}$, if and only if it is a solution to the normal equation $A^TA\vec{x} = A^T\vec{b}$.

□

Example 2. The normal equation, $A^TA\vec{x} = A^T\vec{b}$ can be used to solve for a best fit line of the data in example 1 that we were originally unable to solve. The given data was

X	Y
1	2
2	6
4	1

Therefore,

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 1 \end{pmatrix}$$

$$A^T = \begin{pmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \end{pmatrix}$$

$$\vec{b} = \begin{pmatrix} 2 \\ 6 \\ 1 \end{pmatrix}$$

Substituting these matrices and vector in $A^TA\vec{x} = A^T\vec{b}$ and computing, we get

$$\left(\begin{array}{cc|c} 21 & 7 & 18 \\ 7 & 3 & 9 \end{array}\right) \xrightarrow{RR} \left(\begin{array}{cc|c} 1 & 0 & (19/14) \\ 0 & 1 & -(3/2) \end{array}\right)$$

Notice how the matrix is consistent. The best fit line equation is made by the equation for a line, $\hat{y} = m\hat{x} + b$. Therefore, we get

$$\hat{y} = (19/14)x + -(3/2).$$

And we have solved for the least square solution for $A\vec{x} = \vec{b}$.